Figure S1.

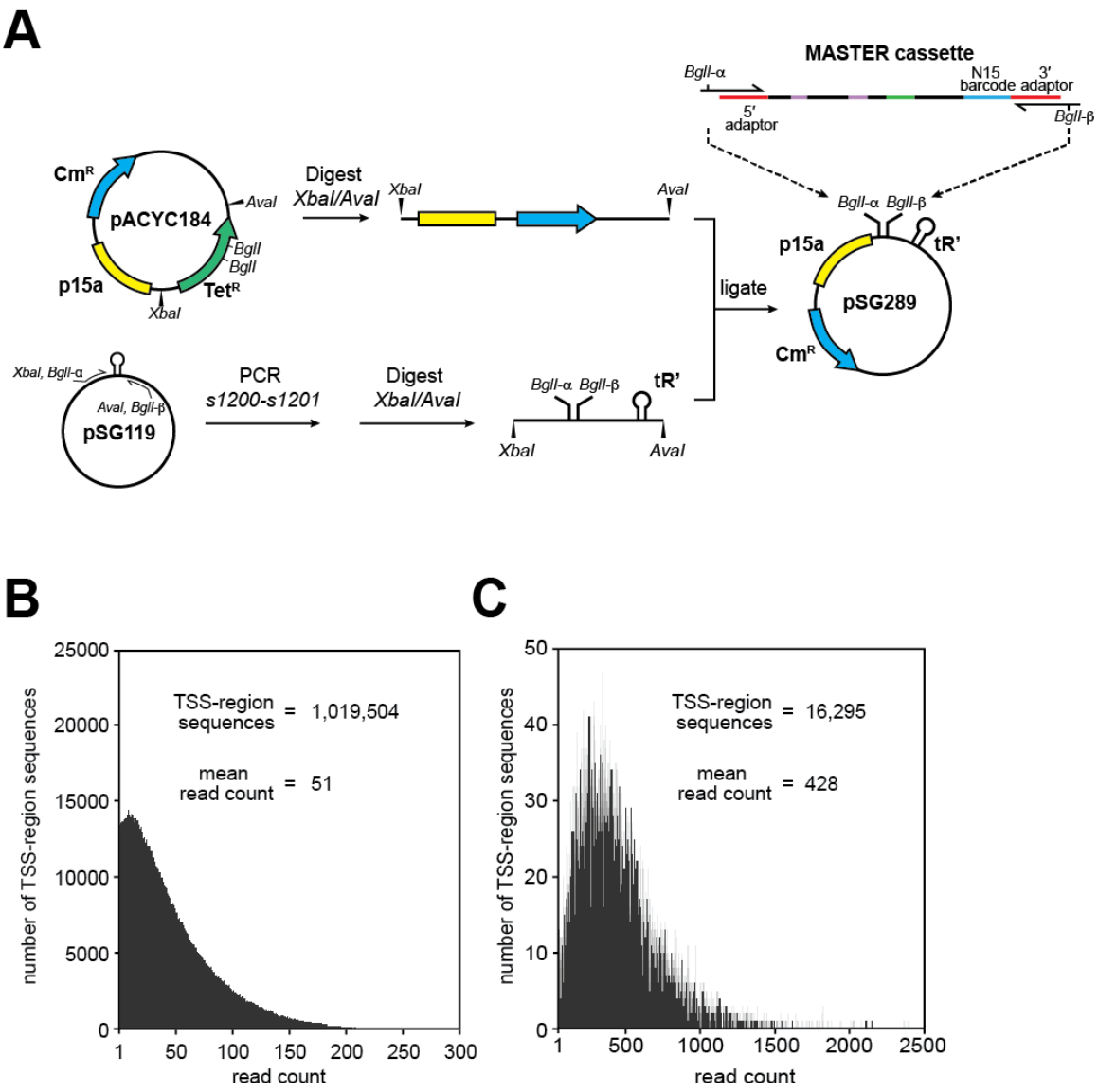

A
MASTER cassette
BglI-α
N15
3'
barcode adaptor
5'
adaptor
BglI-β
CmR
pACYC184
AvaI
BglI
BglI
p15a
TetR
XbaI
Digest
XbaI/AvaI
XbaI
AvaI
BglI-α BglI-β
p15a
tR'
ligate
pSG289
CmR
XbaI, BglI-α
AvaI, BglI-β
pSG119
PCR
s1200-s1201
Digest
XbaI/AvaI
BglI-α BglI-β
tR'
XbaI
AvaI
B
TSS-region sequences = 1,019,504
mean read count = 51
number of TSS-region sequences
read count
C
TSS-region sequences = 16,295
mean read count = 428
number of TSS-region sequences
read count

Figure S1, related to Figure 1. Construction of MASTER libraries and sequence distributions.

A. Construction of vector pSG289. To generate pSG289 pACYC184 was digested with XbaI/AvaI and the fragment containing the p15a origin and chloramphenicol resistance gene was isolated. This DNA fragment was ligated with an XbaI and AvaI digested PCR product that was generated using plasmid pSG119 (Goldman et al., 2011) as template and primers s1200/s1201. The resulting plasmid, pSG289, contains a p15a replication origin, a chloramphenicol resistance gene, and features BglI-α/β recognition sequences upstream of a tR2 transcription terminator. To generate MASTER libraries, DNA fragments are cloned into the BglI-α/β sites of pSG289.

B. Coverage histogram for pMASTER-*lacCONS-N10* library.

C. Coverage histogram for pMASTER-*lacCONS-N7* library.

Figure S2.

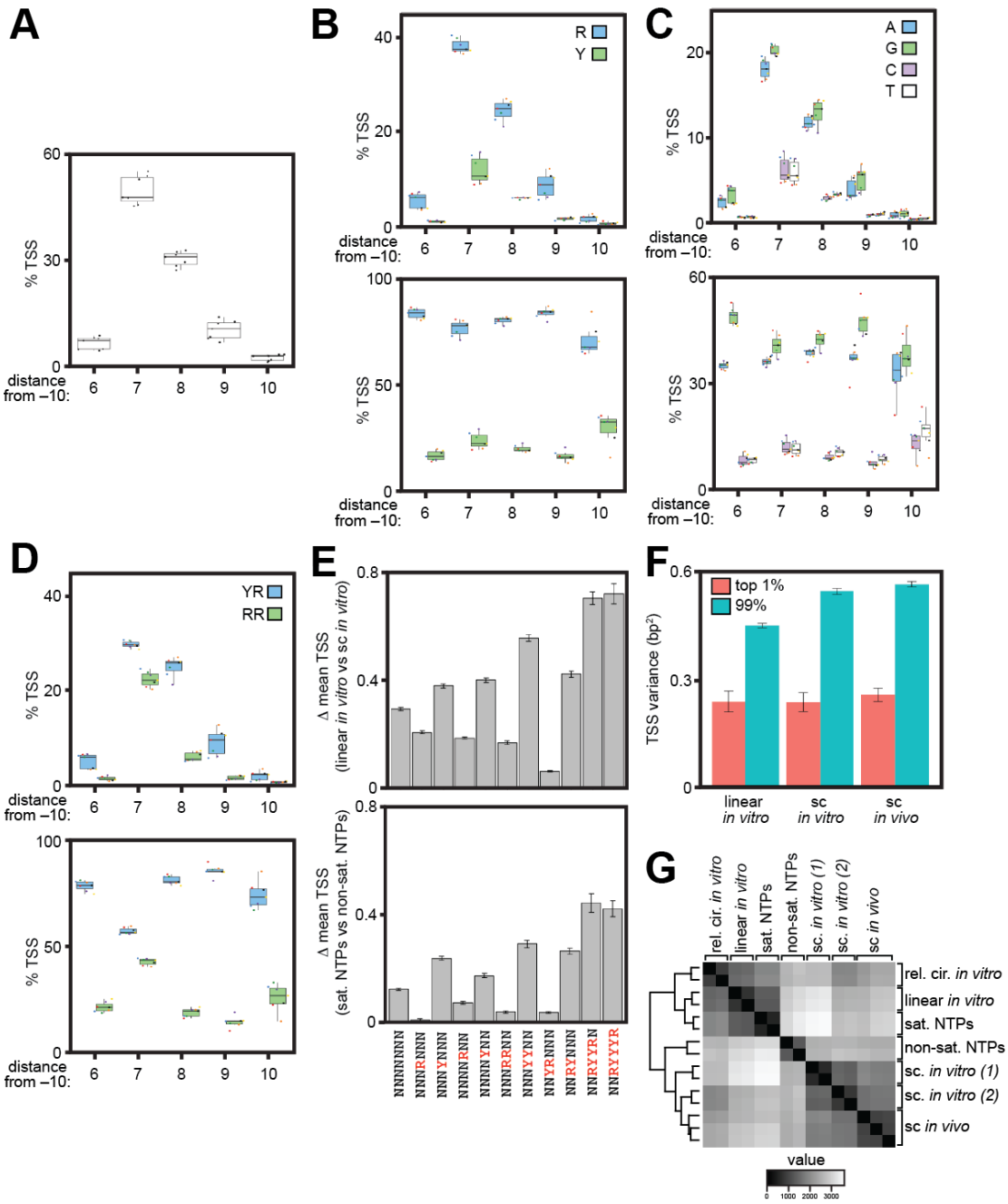

A
B
C
D
E
F
G
% TSS
distance from –10:
6 7 8 9 10
R
Y
A
G
C
T
YR
RR
top 1%
99%
TSS variance (bp²)
linear in vitro
sc in vitro
sc in vivo
Δ mean TSS (linear in vitro vs sc in vitro)
Δ mean TSS (sat. NTPs vs non-sat. NTPs)
NNNNNNN
NNNRNNN
NNNYNNN
NNNNRNN
NNNNYNN
NNNRRNN
NNNYYNN
NNYRNNN
NNRYNNN
NNRYYRN
NNRYYYR
rel. cir. in vitro
linear in vitro
sat. NTPs
non-sat. NTPs
sc. in vitro (1)
sc. in vitro (2)
sc in vivo
value
0 1000 2000 3000

Figure S2, related to Experimental Procedures. Statistical analysis.

A. Boxplot of average % TSS derived from each experimental condition examined in this work. Average % TSS at positions 6-10 was calculated using all TSS-regions with ≥25 matched RNA reads (Tables S1-S7).

B. Boxplot of average % TSS for TSS-regions carrying R or Y at positions 6-10. The graph at the top shows the average % TSS at positions 6-10 calculated using all TSS-regions with ≥25 matched RNA reads (Tables S1-S7). The graph at the bottom shows the % TSS values normalized to a total of 100% for each position. Data points for each data set are depicted in distinct colors.

C. Boxplot of average % TSS for TSS-regions carrying A, G, C or T at positions 6-10. The graph at the top shows the average % TSS at positions 6-10 calculated using all TSS-regions with ≥25 matched RNA reads (Tables S1-S7). The graph at the bottom shows the % TSS values normalized to a total of 100% for each position. Data points for each data set are depicted in distinct colors.

D. Boxplot of average % TSS for TSS-regions carrying $Y_{TSS-1}R_{TSS}$ or $R_{TSS-1}R_{TSS}$ at positions 6-10. The graph at the top shows the average % TSS at positions 6-10 calculated using all TSS-regions with ≥25 matched RNA reads (Tables S1-S7). The graph at the bottom shows the % TSS values normalized to a total of 100% for each position. Data points for each data set are depicted in distinct colors.

E. Effects of topology and NTP concentrations on TSS selection. Height represents the average of the Δmean TSS for the indicated TSS-region sequences. Error bar represents the 95% confidence interval of the average of Δmean TSS, calculated using percentile method based on 10,000 bootstrap replicates. Top graph shows results obtained from comparing the mean TSS values observed on linear and negatively supercoiled templates *in vitro*, bottom graph shows results obtained from comparing the mean TSS values observed at saturating and non-saturating NTP concentrations *in vitro*.

F. Relationship between TSS variance and relative expression. Graph shows the average of TSS variance for TSS-region sequences with the highest 1% relative expression (top 1%) and the variance for TSS-region sequences not in the highest 1% relative expression (99%). Error bar represents the 95% confidence interval of the average of TSS variance, calculated using percentile method based on 10,000 bootstrap replicates.

G. Heat map and hierarchical clustering of the Euclidean distances between the percentages of reads as % TSS distance. Color represents the % TSS distance between the 2 samples. Hierarchical clustering was performed on the % TSS distances between each pair of samples.

Figure S3.

A negatively supercoiled DNA *in vitro*

TSS	%TSS at each position				
	6	7	8	9	10
N	4.3	45.3	32.8	13.9	3.2
R	3.5	36.3	26.8	12.1	2.7
Y	0.9	9.0	6.0	1.8	0.5
A	1.4	16.7	12.6	5.4	1.2
G	2.0	19.7	14.3	6.7	1.5
C	0.5	4.7	2.9	0.8	0.2
T	0.4	4.3	3.1	1.0	0.3

$TSS_{-1}TSS$	%TSS at each position				
	6	7	8	9	10
YR	3.3	29.6	27.0	12.7	3.5
RR	0.8	20.2	7.0	2.0	0.6

B negatively supercoiled DNA *in vivo*

TSS	%TSS at each position				
	6	7	8	9	10
N	4.7	47.6	32.3	12.0	2.9
R	3.8	37.0	26.2	10.2	2.0
Y	0.9	10.6	6.1	1.8	0.9
A	1.6	17.4	12.1	4.4	1.1
G	2.2	19.7	14.2	5.8	0.9
C	0.5	5.4	2.9	0.8	0.4
T	0.4	5.2	3.2	1.0	0.5

$TSS_{-1}TSS$	%TSS at each position				
	6	7	8	9	10
YR	3.4	28.7	26.0	10.5	2.2
RR	1.2	22.1	7.3	1.9	0.8

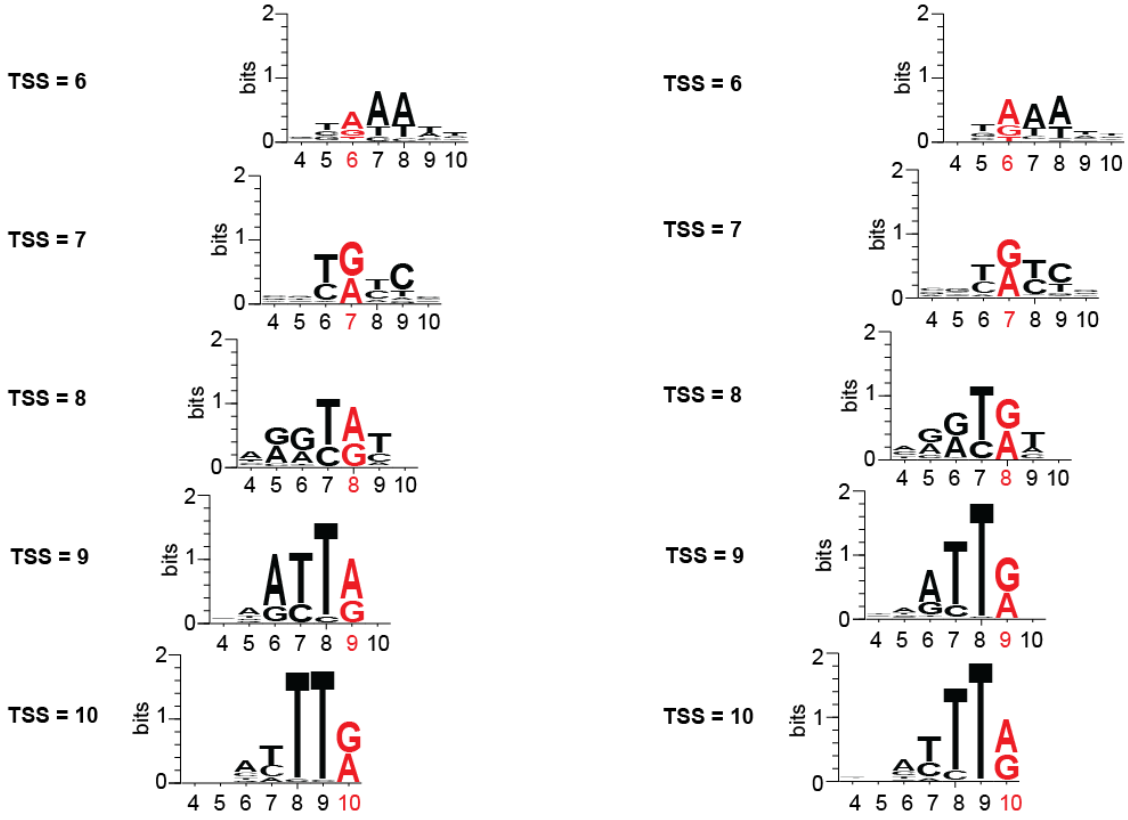
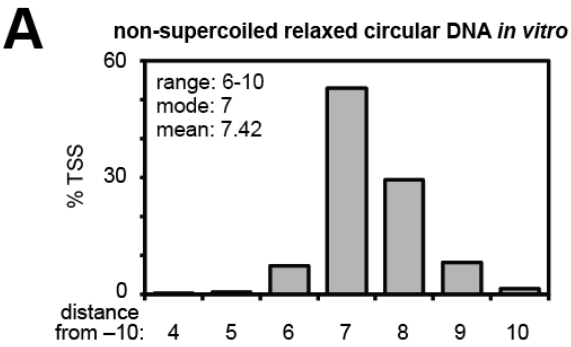
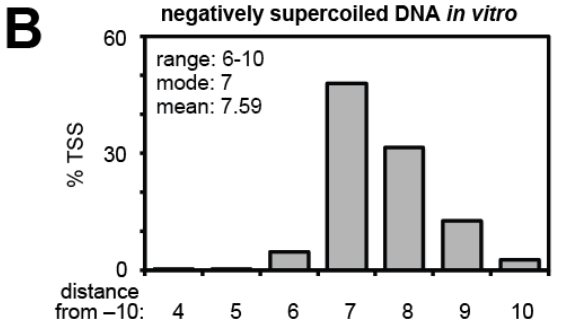


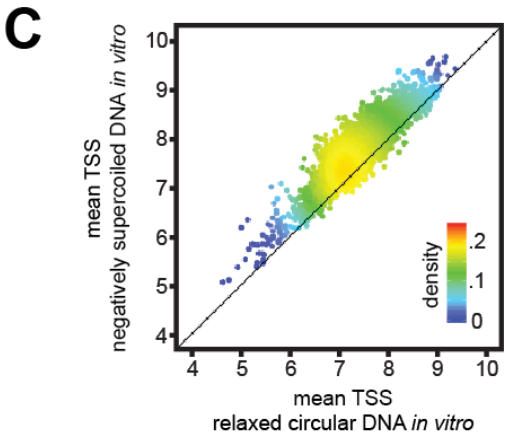
Figure S3, related to Figure 2. TSS selection on negatively supercoiled DNA templates *in vitro* and *in vivo*.

Top of each panel shows sequence determinants for TSS selection. Table lists the amount of the total %TSS at positions 6-10 derived from TSS-regions carrying (i) R or Y at the indicated TSS position, (ii) A, G, C, or T at the indicated TSS position, or (iii) $Y_{TSS-1}R_{TSS}$ or $R_{TSS-1}R_{TSS}$ at the indicated TSS position. Bottom panels show sequence preferences for TSS selection. Sequence logos for the 162 TSS-region sequences (top 1%) with the highest value for %TSS at positions 6-10. Red bases indicate the TSS position. (A) Results obtained with negatively supercoiled DNA templates *in vitro*. (B) Results obtained with negatively supercoiled DNA templates *in vivo*.

Figure S4.







D

sequence	mean TSS: relaxed circular *in vitro*	mean TSS: sc *in vitro*	Δ mean TSS: relaxed circular *in vitro* vs sc *in vitro*
NNNNNNN	7.42	7.59	0.18
NNNRNNN	7.14	7.27	0.13
NNNYNNN	7.70	7.92	0.22
NNNNRNN	7.49	7.60	0.11
NNNNYNN	7.35	7.59	0.25
NNNRRNN	7.14	7.25	0.10
NNNYYNN	7.56	7.90	0.35
NNYRNNN	7.14	7.20	0.06
NNRYNNN	7.80	8.06	0.26
NNRYYRN	8.06	8.45	0.39
NNRYYYR	7.62	8.07	0.45

E

	mean TSS
relaxed circular DNA *in vitro*	7.16
negatively supercoiled DNA *in vitro*	8.17

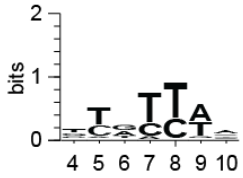


Figure S4, related to Figure 3. Effects of DNA topology on TSS selection

A. TSS-distribution histogram for non-supercoiled relaxed circular DNA templates *in vitro*. Average %TSS at positions 4-10 for all TSS-regions in the *lacCONS-N7* library with ≥25 matched RNA reads (Table S6). Experiments were performed in the presence of 1 mM NTPs.

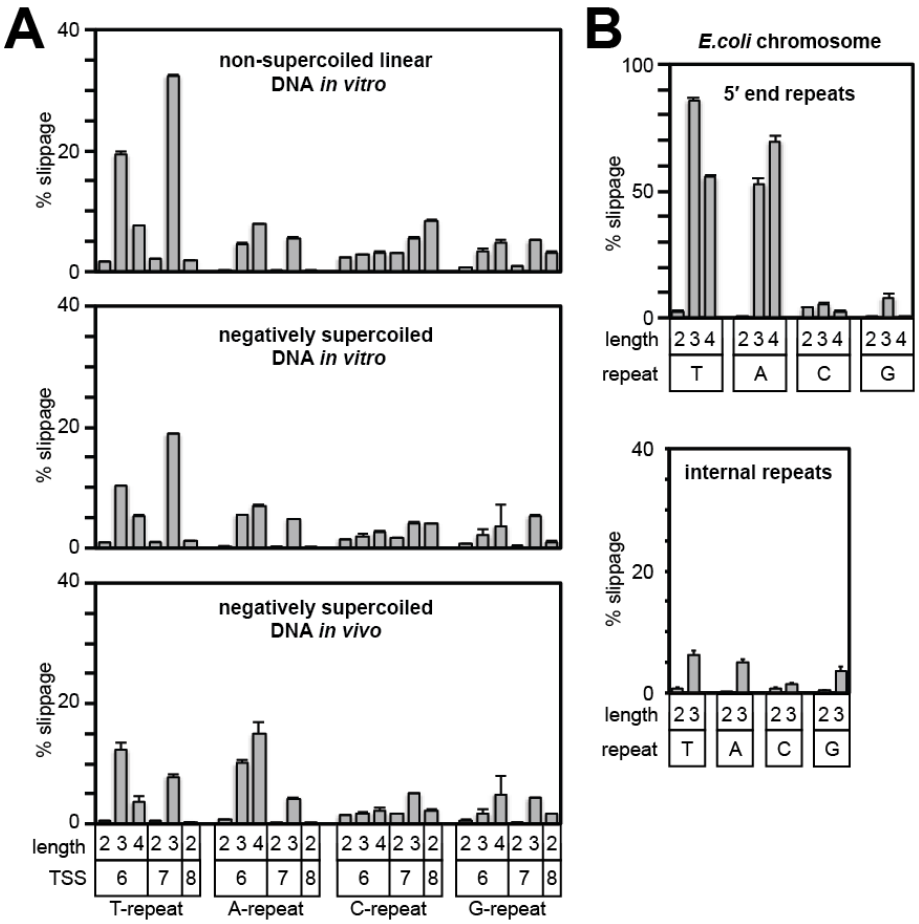
B. TSS-distribution histogram for negatively supercoiled DNA templates *in vitro*. Average %TSS at positions 4-10 for all TSS-regions in the *lacCONS-N7* library with ≥25 matched RNA reads (Table S7). Experiments were performed in the presence of 1 mM NTPs. These reactions are distinct from those presented in Figure 3, and were performed in parallel with those performed using non-supercoiled relaxed circular DNA templates.

C. Plot of the mean TSS observed with negatively supercoiled DNA and non-supercoiled relaxed circular DNA for individual TSS-region sequences.

D. Average of the mean TSS values observed for the indicated TSS-region sequences with non-supercoiled relaxed circular DNA and negatively supercoiled DNA. The differences between the values observed with non-supercoiled relaxed circular DNA and negatively supercoiled DNA (Δ mean TSS) are also listed.

E. Sequence preferences for topology dependent effects on TSS selection. Sequence logo and the average mean TSS values for the 162 TSS-region sequences (top 1%) with the highest values of Δ mean TSS in the comparison of non-supercoiled relaxed circular DNA and negatively supercoiled DNA.

Figure S5.



C

pattern	dinucleotide											
	TG	TA	CG	AG	TC	CA	GT	AT	CT	GC	AC	GA
ab	1.9	0.5	1.4	0.7	1.4	0.2	1.3	1.1	0.5	0.3	0.6	0.1
bab	30.6	10.8	8.7	3.1	3.2	3.6	1.4	1.0	0.7	0.3	0.6	0.4
abab	50.7	21.7	15.6	12.7	5.9	6.1	1.2	1.5	1.2	0.4	1.2	0.7
aba	1.7	0.7	1.4	1.2	3.6	0.4	4.6	4.3	2.3	1.4	1.9	0.4
baba	23.1	8.8	8.8	3.8	6.3	5.1	4.8	4.0	2.9	1.4	1.9	1.4
ababa	37.0	14.1	14.9	11.2	8.4	8.8	4.1	5.0	4.2	1.3	2.5	1.9
ab	0.7	0.3	0.5	0.3	1.2	0.1	0.5	0.8	1.0	0.1	0.6	0.0
bab	13.1	5.3	3.6	1.5	2.7	2.6	0.5	0.9	1.4	0.2	1.1	0.2
abab	27.0	11.2	5.9	5.8	3.7	5.4	0.6	1.5	2.3	0.3	1.5	0.4
aba	0.7	0.2	0.7	0.7	3.1	0.3	2.2	3.3	2.7	1.0	2.1	0.2
baba	9.8	6.2	5.0	2.5	6.5	3.5	2.4	4.0	4.0	1.5	3.5	1.0
ababa	18.5	8.4	8.0	6.4	7.6	6.9	2.5	5.9	6.3	1.5	4.2	1.2
ab	0.5	0.4	0.5	0.3	1.4	0.2	0.4	0.8	0.8	0.1	0.8	0.1
bab	7.5	5.5	4.5	1.5	2.7	3.5	0.3	0.9	1.0	0.1	0.9	0.3
abab	15.1	10.0	6.6	5.4	3.9	5.7	0.3	1.5	1.7	0.1	1.4	0.7
aba	0.7	0.6	0.5	1.1	3.6	0.4	1.9	3.3	1.9	1.1	2.6	0.3
baba	5.5	8.5	4.2	4.5	5.9	5.2	1.6	4.2	2.4	0.5	2.9	1.6
ababa	9.2	12.0	5.6	10.9	6.3	7.3	1.3	5.3	3.5	0.5	3.5	2.1
ab	0.1	0.0	0.0	0.4	2.1	0.2	0.4	1.1	0.4	0.3	2.2	0.1

non-supercoiled linear DNA *in vitro*

negatively supercoiled DNA *in vitro*

negatively supercoiled DNA *in vivo*

E.coli chromosome

Figure S5, related to Figure 5. Analysis of productive slippage synthesis *in vitro* and *in vivo*
A. Productive slippage synthesis from TSS-region sequences with a 2, 3, or 4 base pair internal homopolymeric repeat sequence. Graphs show the % slippage (mean + SEM) observed for TSS-region sequences carrying 2, 3, or 4 consecutive T, A, C, or G bases beginning one base pair downstream of the indicated TSS position.
B. Productive slippage synthesis from *E. coli* chromosomal promoters with TSS-region sequences containing homopolymeric repeats [identified in (Thomason et al., 2015)]. Graph at the top shows the % slippage (mean + SEM) observed for TSS-region sequences carrying 2, 3, or 4 consecutive T, A, C, or G bases beginning at the TSS (5′ end repeats). Graph at the bottom shows the % slippage observed for TSS-region sequences carrying 2 or 3 consecutive T, A, C, or G bases beginning at the position downstream of the TSS (internal repeats). Results represent the reanalysis of the raw sequencing data from (Thomason et al., 2015) for transcripts generated by productive slippage synthesis.
C. Productive slippage synthesis from TSS-region sequences with non-repeat dinucleotide sequences. Table reports the % slippage values observed for TSS-region sequences carrying the indicated dinucleotide pattern (ab) at position 7. For results of MASTER analysis the box has been drawn around positions 7 and 8 of each sequence pattern. The *E. coli* chromosomal promoter analysis at the bottom of the Table was performed by reanalyzing the raw sequencing data from (Thomason et al., 2015) for transcripts generated by productive slippage synthesis. For the *E. coli* chromosomal promoter analysis the box has been drawn around bases representing the TSS and TSS_{+1} [as identified in (Thomason et al., 2015)].

Figure S6.

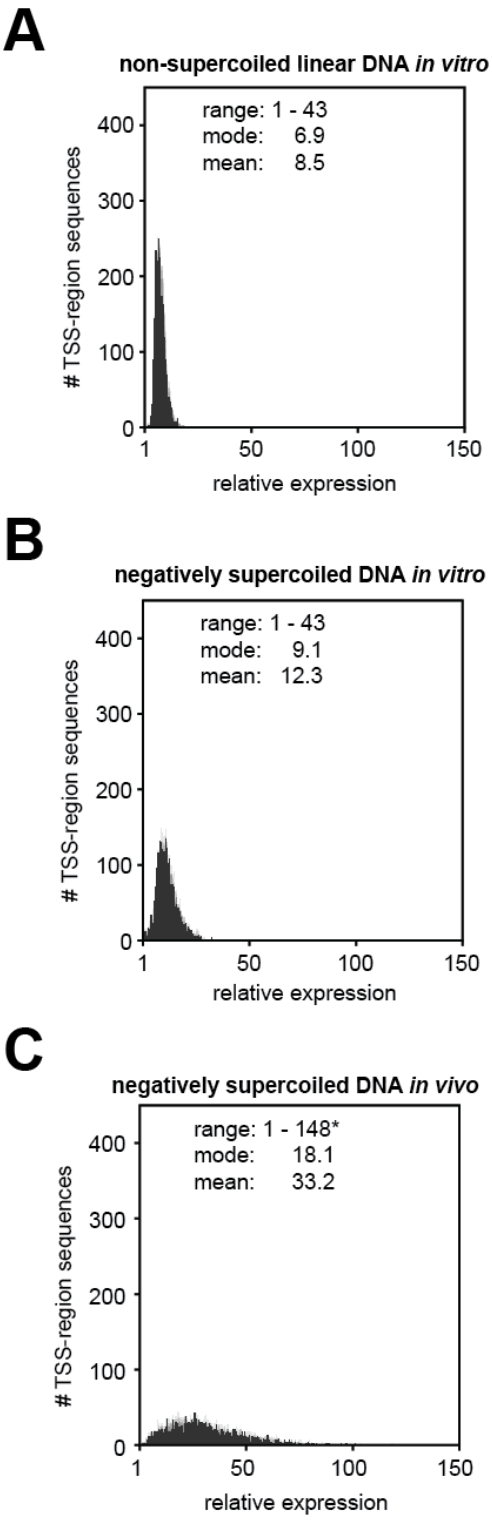

A
non-supercoiled linear DNA *in vitro*
range: 1 - 43
mode: 6.9
mean: 8.5
TSS-region sequences
relative expression
B
negatively supercoiled DNA *in vitro*
range: 1 - 43
mode: 9.1
mean: 12.3
TSS-region sequences
relative expression
C
negatively supercoiled DNA *in vivo*
range: 1 - 148*
mode: 18.1
mean: 33.2
TSS-region sequences
relative expression

Figure S6, related to Figure 6. Effects of DNA topology on transcript yields

A. Relative expression histograms for experiments performed using non-supercoiled linear DNA *in vitro*.

B. Relative expression histograms for experiments performed using negatively supercoiled DNA *in vitro*.

C. Relative expression histograms for experiments performed using negatively supercoiled DNA *in vivo*. Relative expression for TSS-region sequences with ≥25 total RNA reads and for which the number of DNA templates was not in the top or bottom 10% (Tables S1, S2, and S3). Asterisk in panel C indicates that relative expression values for the following 14 TSS-region sequences which are above 150 are not shown (ACCGTTC, CGCGTTC, CTGGCTC, CTGGTTC, GCGTGTA, GCGTGTC, GCGTGTG, GCTGTTC, GCTGTTG, GCTTGTG, GGGGTTC, TATTGGC, TCGGGTG, TTCGTTC).

SUPPLEMENTAL TABLE LIST

Table S1, related to Figures 2, 3, 5, 7, S2, S5, S6. Transcription output for experiments performed *in vitro* using a non-supercoiled linear template (at 1 mM NTPs).

Table S2, related to Figures 3, 5, 7, S2, S3, S5, S6. Transcription output for experiments performed *in vitro* using a negatively supercoiled template (at 1 mM NTPs).

Table S3, related to Figures 3, 5, 7, S2, S3, S5, S6. Transcription output for experiments performed *in vivo* using a negatively supercoiled template.

Table S4, related to Figures 4, 6, S2. Transcription output for experiments performed *in vitro* at saturating NTPs (2.5 mM:Mg2+) using a non-supercoiled linear template.

Table S5, related to Figures 4, 6, S2. Transcription output for experiments performed *in vitro* at non-saturating NTPs (0.1 mM) using a non-supercoiled linear template.

Table S6, related to Figures 2, S4, S2. Transcription output for experiments performed *in vitro* using a non-supercoiled relaxed circular template (at 1 mM NTPs).

Table S7, related to Figures 2, S4, S2. Transcription output for experiments performed *in vitro* using a negatively supercoiled template (at 1 mM NTPs), second set.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Construction of MASTER plasmid libraries

Overview of strategy: Our strategy for constructing MASTER libraries comprises five steps (i) synthesize a DNA oligo containing two randomized regions (one serving as the region of interest and a second, longer region serving as a barcode); (ii) convert the DNA oligo to double-stranded DNA by PCR (during this PCR step restriction sites are introduced that enable directional cloning of the PCR product into a plasmid); (iii) insert the PCR product into a plasmid backbone (in the new recombinant plasmid, the insert is positioned upstream of an intrinsic transcription terminator); (iv) transform recombinant plasmids into cells (the number of transformants obtained at this step determines the complexity of the library); (v) isolate DNA from ~0.5-10 million individual transformants.

Generation of a parent vector (pSG289) for use in MASTER library construction (Figure S1A): To avoid loss of sequences in the region of interest during cloning, we took advantage of the sequence recognition properties of BglI. The BglI recognition sequence, GCCnnnnnGGC, cannot be inadvertantly created by any combination of the randomized region of interest in the MASTER-*lacCONS-N7* library and its flanking sequences. Furthermore, the five variable bps in the middle of the BglI recognition sequence facilitate directional cloning by use of two variant BglI sites that yield incompatible overhangs after cleavage (BglI-α, GCCtgaccGGC; BglI-β, GCCccagcGGC).

To generate pSG289 we digested pACYC184 with XbaI/AvaI and isolated the fragment containing the p15a origin and chloramphenicol resistance gene. Next, we generated a PCR product using plasmid pSG119 (Goldman et al., 2011) as template and oligonucleotides s1200/s1201 (see below) as primers. This PCR product contains, in order: an XbaI site, a BglI-α site, a BglI-β site, the tR2 terminator (from pSG119), and an AvaI site. The PCR product was digested with XbaI/AvaI and ligated with the p15a origin-containing fragment from pACYC184 to generate pSG289. Oligonucleotides: s1200, upstream pSG119 primer, 5'-taaatatctagagcctgaccggcattatagccccagcggcCGTCTTCAAGAgcTCATGGATCC-3'; s1201, downstream pSG119 primer, 5'-tataatcccgagGTCGACATCCAAATAAAAAGGCCTGC-3'. Sequences in upper case are complementary to the template.

Preparation of PCR product containing lacCONS-N7 and lacCONS-N10 libraries:
A PCR cocktail containing either 50 nM ssDNA oligo JB1 or oligonucleotide s1250 as template, 5 µM primer s1219, 5 µM primer s1220 and 1X Q5 DNA polymerase master mix (NEB) was prepared in a volume of 20 µl in a thin-walled 200 µl PCR tube. The reaction was thermally cycled: 98°C x 30s; 35 cycles (98°C x 2s, 64°C x 5s, 72°C x 5s); 72°C x 60s. Afterwards, the PCR mixture was separated by agarose electrophoresis and the product band was extracted and purified using a gel extraction kit (Qiagen). Oligonucleotides: JB1, *lacCONS-N7* ssDNA template, 5'-
GTTCAGAGTTCTACAGTCCGACGATCaggc**TTGACA**ctttatgcttcggctcg**TATAAT**gtgNNNN
NNNgtgagcggataacaatNNNNNNNNNNNNNNNTGGAATTCTCGGGTGCCAAGG-3'
S1250, *lacCONS-N10* ssDNA template, 5'-
GTTCAGAGTTCTACAGTCCGACGATCgcggccgcaggc**TTGACA**ctttatgcttcggctcg**TATAAT**N
NNNNNNNNNgtgagcggataacaatNNNNNNNNNNNNNNNcctgcaggTGGAATTCTCGGGTGCC
AAGG-3'
(for JB1 and s1250 the -35 and -10 promoter elements are indicated in bold uppercase and randomized bases are denoted "N", non-bold uppercase letters denote regions in the dsDNA product complementary to PCR primers; s1219, upstream PCR primer, 5'-tataatgcctgaccggcGTTCAGAGTTCTACAGTCCGACGATC-3'; s1220, downstream PCR primer, 5'-aattaagccgctggggcCCTTGGCACCCGAGAATTCC-3'.

Cloning of PCR product containing lacCONS-N7 and lacCONS-N10 libraries into pSG289
Restriction cocktails containing 20U of BglI (NEB), 1X NEB3.1 buffer and either 500 ng of purified PCR product or 200 ng of pSG289 vector were prepared in a total volume of 50 µl. PCR digests were incubated at 37°C for 2 h. Vector digests were incubated at 37°C for 90 min at which time 5.5 µl of 10X Antarctic Phosphatase buffer and 20U of Antarctic Phosphatase were added and the reaction was mixed and returned to 37°C for an additional 30 min. After digestion, reactions were stopped by addition of 5 volumes of Qiagen solution PB and purified over QIAquick reaction cleanup columns according to the standard protocol and eluted in 30 µl of water.

A 100 µl ligation reaction containing ~25 ng of digested PCR product, ~50 ng of digested pSG289, 1X T4 DNA ligase buffer (NEB), and 2,000U T4 DNA ligase (NEB) was prepared on ice and then incubated for 16 hours at 16°C. The ligation reaction was extracted with phenol:chloroform:IAA (25:24:1) (pH 8), and DNA was recovered by ethanol precipitation. To

ensure complete removal of all phenol and trace salts, pellets were washed three times with 1 ml portions of 75% ethanol prior to resuspension in water to a concentration of 1 µg/µl. To generate the plasmid library, 1 µg of the ligation reaction was mixed with 25 µl of commercially prepared electrocompetent NEB 5-alpha cells (New England Biolabs) in a chilled 0.1 mm gap electroporation cuvette and pulsed at 1.8 kV using a MicroPulser (Bio-Rad), cells were immediately recovered by addition of 975 µl SOC broth (NEB) directly to the cuvette, then transferred to 1.7 ml tubes and gently inverted for 1 hour at 37°C. A 10 µl aliquot of the cell suspension was serially diluted and plated for CFU determination while the remaining ~990 µl was inoculated directly into 50 ml of LB media [10 g Bacto tryptone (Becton Dickinson & Co), 5 g Bacto yeast extract (Becton Dickinson & Co) and 10 g NaCl per liter] containing chloramphenicol (25 µg/ml) in a 250 ml DeLong flask with Morton-style closure (Bellco) and shaken at 220 RPM at 37°C for 16 hours at 50% relative humidity. The transformation procedure was repeated to obtain the desired number of transformants. In the case of the pMASTER-*lacCONS-N7* library we obtained ~500,000-1,000,000 individual transformants. In the case of the pMASTER-*lacCONS-N10* library we obtained ~5,000,000 individual transformants. (We note that transformation of the ligation mixture into bacterial cells limits the number of possible 15-bp barcodes in the library to the total number of transformants obtained in this step.) Libraries generated during this procedure were isolated from the 50 ml cell suspension using a Plasmid Plus Midi-prep kit (Qiagen).

Generation of DNA templates for in vitro transcription assays

Negatively supercoiled DNA templates: ~50 ng of pMASTER-*lacCONS-N7* was introduced into commercially prepared electrocompetent DH10B-T1^{R} cells (Life Technologies) to obtain a 50 ml overnight cultures representing cells derived from ~20 million unique transformants (as estimated by CFU determination). The 50 ml overnight cell culture was used to inoculate 1 L of LB media containing chloramphenicol in a 2.8 L baffled Fernbach flask and shaken at 175 RPM at 37°C for ~4 h to reach OD_{600} ~0.8 whereupon spectinomycin was added at 100 µg/ml to increase the copy number of pMASTER-*lacCONS-N7* (by relieving the repression of the plasmid origin of replication). Cell cultures were shaken for additional 16 h, divided equally among six 500 ml centrifuge bottles, and cells were collected by centrifugation at 13000 x *g* at 4°C for 15 min. Supernatants were removed and plasmid DNA was isolated using a Plasmid Mega Kit (Qiagen). This stock of the pMASTER-*lacCONS-N7* library was used for two purposes: (i) as a template in emulsion PCR (ePCR) to generate a linear DNA template for *in vitro* transcription experiments

and (ii) as a template for *in vitro* transcription reactions performed using negatively supercoiled templates.

Relaxed circular DNA templates: Relaxed plasmids (ΔLk=0) were generated starting from positively supercoiled plasmids (Sc+). A Sc+ plasmid was obtained by incubating purified negatively supercoiled plasmids (Sc-) with *Pyrobaculum calidifontis* reverse gyrase (Jamroze et al., 2014) for 10 min at 90°C at 3-fold molar excess of protein to DNA in the presence of 1 mM ATP. After incubation, Sc+ plasmids were purified by phenol extraction and ethanol precipitation and checked by two-dimensional gel electrophoresis in 1.2% agarose, 1X TBE (Napoli et al., 2005). In order to obtain relaxed plasmids, purified Sc+ plasmids were incubated with wg-Topoisomerase I (Promega) at protein/DNA ratio of 20U/µg for 30 min at 37 °C in 50 mM Tris-HCl (pH 7.5), 50 mM NaCl, 0.1 mM EDTA, 1 mM DTT. After incubation, the reaction was stopped with 2% of SDS and DNA was purified by phenol extraction and ethanol precipitation. To check the reaction, an aliquot of the sample was run on 1.2% agarose gel in 1X TBE. Gels were stained with ethidium bromide (1 µg/ml) and destained in deionized water.

Non-supercoiled linear DNA templates. The pMASTER-*lacCONS-N7* plasmid isolated above was diluted to ~10^9 molecules/µl. 1µl of diluted DNA was amplified by ePCR using a Micellula DNA Emulsion and Purification Kit (Chimerx) in detergent-free Phusion HF reaction buffer containing 5 µg/ml BSA, 0.4 mM dNTPs, 0.5 µM Illumina RP1 primer, 0.5 µM Illumina RPI1 primer and 0.04 U/µl Phusion HF polymerase (Thermo Scientific). ePCR reactions were performed with an initial denaturation step of 10 s at 95°C, amplification for 30 cycles (denaturation for 5 s at 95°C, annealing for 5 s at 60°C and extension for 15 s at 72°C), and a final extension for 5 min at 72°C. The emulsion was broken and DNA was purified according to the manufacturer's recommendations. The DNA was recovered by ethanol precipitation, the pellet was washed twice with 80% cold ethanol, and resuspended in 20 µl of nuclease-free water. Oligos: Illumina RP1, 5'-AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGA-3' Illumina RPI1, 5'-CAAGCAGAAGACGGCATACGAGAT**CGTGAT**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA-3' (where the index sequence is in bold).

In vitro transcription

10 nM of template DNA (negatively supercoiled, relaxed or linear) was mixed with 50 nM RNAP holoenzyme [prepared as described in (Vvedenskaya et al., 2014)] in transcription buffer containing 50 mM Tris HCl (pH 8.0), 10 mM $MgCl_2$, 0.01 mg/ml BSA, 100 mM KCl, 5% glycerol, 10 mM DTT, 0.4U/µl RNase OUT. Open complexes were allowed to form by incubation at 37°C for 10 minutes. A single round of transcription was initiated by addition of a mixture of NTPs (to a final concentration of 10, 1, or 0.1 mM) and heparin (to a final concentration of 0.1 mg/ml). After 15 minutes, the reactions were stopped by addition of 0.5 M EDTA (pH 8) to a final concentration of 10 mM. Nucleic acids were recovered by ethanol precipitation, washed twice with 80% cold ethanol, and resuspended in 30 µl of nuclease-free water.

Purification of RNAs generated in vitro

Reactions performed *in vitro* using a linear DNA template were treated with 2U TURBO DNase (Life Technologies) at 37°C for 1 h, mixed with an equal volume of 2x RNA loading dye (95% deionized formamide, 18 mM EDTA, 0.25% SDS, xylene cyanol, bromophenol blue, amaranth), and separated by electrophoresis on 10% 7M urea slab gels (equilibrated and run in 1x TBE). The gel was stained with SYBR Gold nucleic acid gel stain (Life Technologies), bands were visualized on a UV transilluminator, and RNA transcripts ~100 nt in size were excised from the gel. The excised gel slice was crushed and incubated in 300 µl of 0.3 M NaCl in 1x TE buffer at 70°C for 10 min. Eluted RNAs were separated from crushed gel fragments using a Spin-X column (Corning). After the first elution, the crushed gel fragments were collected and the elution procedure was repeated, nucleic acids were collected, pooled with the first elution, isolated by ethanol precipitation, and resuspended in 25 µl of RNase-free water.

Reactions performed *in vitro* using a plasmid-borne template were mixed with an equal volume of 2x RNA loading dye and separated by electrophoresis on 10% 7M urea slab gels (equilibrated and run in 1x TBE). Nucleic acids were visualized as above and RNA transcripts ~140 nt in size were excised from the gel. RNA was eluted from gel as described above, isolated by ethanol precipitation and resuspended in 25 µl of RNase-free water.

Generation and isolation of RNAs generated in vivo

~50 ng of pMASTER-*lacCONS-N7* was introduced into DH10B-$T1^R$ cells to obtain a 50 ml overnight cell culture that contained ~20 million individual transformants. 0.5 ml of this culture was used to inoculate 50 ml of LB media containing chloramphenicol (25 µg/ml) in a 250 ml

DeLong flask and shaken at 210 RPM at 37°C to mid-exponential phase (OD600 ~0.5). 2 ml aliquots of the cell suspensions were placed in 2 ml tubes (BioExcell) and cells were collected by centrifugation (1 min, 21,000 x *g* at room temperature). Supernatants were removed and cell pellets were rapidly frozen on dry ice and stored at -80°C. RNA isolated from these cell pellets (as described below) was used for 5′ RNA-seq. pMASTER-*lacCONS-N7* plasmid DNA was also isolated from these cells using a Plasmid Mini-prep kit (Qiagen). Plasmid DNA was used as template in ePCR reactions to generate a product that was sequenced to assign barcodes (see below).

RNA isolation and purification was performed essentially as in (Vvedenskaya et al., 2015). 600 μl of TRI Reagent solution (Molecular Research Center) was added to a cell pellet collected from 2 ml of cell suspension. Samples were incubated at 70°C for 10 min, centrifuged (10 min, 21,000 x *g*) to remove insoluble material. The supernatant was transferred to a fresh tube, ethanol was added to a final concentration of 60.5%, and the mixture was applied to a Direct-zol spin column (Zymo Research). DNase I treatment was performed on-column according to the manufacturer's recommendations. RNA was eluted from the column with 3 sequential portions of 30 μl nuclease-free water that had been heated to 70°C. To remove residual DNA, eluted RNAs were treated with 2U TURBO DNase (Life Technologies) at 37°C for 1 h. Following DNase treatment, samples were extracted with acid phenol:chloroform, RNA was recovered by ethanol precipitation and resuspended in RNase-free water. A MICROBExpress Kit (Life Technologies) was used to deplete rRNAs from 9 μg of TURBO DNase-treated RNA. The rRNA-depleted RNAs were isolated by ethanol precipitation and resuspended in 30 μl of RNase-free water.

5′ RNA-seq

5′ RNA-seq: enzymatic treatments for analysis of 5′ triphosphate RNA. For RNA isolated from *in vitro* transcription reactions, ~100 ng of RNA was treated with 20 U 5′-RNA Polyphosphatase. Samples were extracted with acid phenol:chloroform (pH 4.5) and RNAs were recovered by ethanol precipitation and resuspended in 10 μl RNase-free water.

For RNA isolated from cells, 2.5 μg of rRNA-depleted RNA was treated with 1 U Terminator 5′-Phosphate-Dependent Exonuclease (Epicentre) to remove 5′ monophosphate RNA, samples were extracted with acid phenol:chloroform, RNA was recovered by ethanol precipitation, and resuspended in RNase-free water. Next, to convert 5′ triphosphate RNA to 5′ monophosphate

RNA, samples were treated with 20 U 5′-RNA Polyphosphatase as described in (Goldman et al., 2011), extracted with acid phenol:chloroform, RNA was recovered by ethanol precipitation, and resuspended in 10 µl RNase-free water.

5′ RNA-seq: ligation of adaptor to 5′ end of RNAs. RNA prepared as described above was combined with PEG 8000 (10% final concentration), oligo s1206 (1 pmol/µl final concentration), ATP (1mM final concentration), 40 U RNase OUT, 1x T4 RNA ligase 1 reaction buffer (NEB) and 10 U of T4 RNA ligase 1 (NEB) in a total volume of 30 µl. The mixture was incubated at 16°C for 16 h. Oligo: s1206, 5'-GUUCAGAGUUCUACAGUCCGACGAUCNNNNNNNNNNNNNNN-3' (all bases are RNA)

5′ RNA-seq: size selection of adaptor ligated RNA. Size selection of RNA ligated to 5′ adaptor was performed as described in (Vvedenskaya et al., 2015). Adaptor-ligated RNAs were mixed with equal volume of 2x RNA loading dye and were separated by electrophoresis on 10% 7M urea slab gels (equilibrated and run in 1x TBE). The gel was stained with SYBR Gold nucleic acid gel stain, bands visualized with UV transillumination, and species ranging from ~80 to ~300 nt were excised from the gel. RNAs were eluted from gel using the procedure described above, isolated by ethanol precipitation and resuspended in 10 µl of nuclease-free water.

5′ RNA-seq: cDNA synthesis. 10 µl of gel-eluted RNAs were mixed with 0.3 µl of s128 oligonucleotide (100 pmol/ µl) and incubated at 65°C for 5 min then cooled to 4°C. 9.7 µl of a cocktail containing 4 µl of 5x First-Strand buffer (Life Technologies), 1 µl of 10 mM dNTP mix, 1 µl of 100 mM DTT, 1 µl (40U) RNase OUT, 1 µl (200U) of SuperScript III Reverse Transcriptase (Life Technologies) and 1.7 µl of nuclease-free H2O were added to the RNA/oligo mixture. The reactions were incubated in a thermal cycler with a heated lid at 25°C for 5 min, followed by 55°C for 60 min and 70°C for 15 min. Reactions were cooled to room temperature, 10U of RNase H (Life Technologies) were added, and the reactions were incubated at 37°C for 20 min. Oligo: s128, Illumina RT primer, 5'-CCTTGGCACCCGAGAATTCC-3'.

5′ RNA-seq: size selection of cDNAs. An equal volume of 2x RNA loading dye was added and nucleic acids were separated by electrophoresis on 10% 7M urea slab gels (equilibrated and run in 1x TBE). The gel was stained with SYBR Gold nucleic acid gel stain and ~80 to ~150 nt species were excised from the gel. cDNAs were recovered from gel using the procedure described

above and resuspended in 13 µl of nuclease-free water.

5′ RNA-seq: amplification of cDNAs. To amplify cDNA, 5 µl of gel-isolated cDNA were added to the mixture containing 1x Phusion HF reaction buffer, 0.2 mM dNTPs, 0.25 µM Illumina RP1 primer, 0.25 µM Illumina index primer and 0.02 U/µl Phusion HF polymerase. PCR was performed with an initial denaturation step of 30 s at 98°C, amplification for 11 cycles (denaturation for 10 s at 98°C, annealing for 20 s at 62°C and extension for 10 s at 72°C), and a final extension for 5 min at 72°C.

Purification of cDNAs. Amplified cDNAs were isolated by electrophoresis using a non-denaturing 10% slab gel (equilibrated and run in 1x TBE). The gel was stained with SYBR Gold nucleic acid gel stain and species around ~170 bp were excised from the gel. cDNA were eluted from gel with 600 µl of 0.3M NaCl in 1xTE buffer at 37° C for 2 h, precipitated and resuspended in 13 µl of nuclease-free water.

High-throughput sequencing. Barcoded libraries were pooled and sequenced on an Illumina HiSeq 2500 platform in rapid mode. To avoid the potentially confounding effects of mis-annealing of the default pooled Illumina sequencing primer cocktail to the randomized sequences of the libraries, custom primer s1115 (5'-CTACACGTTCAGAGTTCTACAGTCCGACGATC-3') was used.

Experiments performed *in vitro* were performed in duplicate. Experiments performed *in vivo* were performed in triplicate. Sample serial numbers for 5′ RNA seq samples are listed below:
(i) VV864 and VV865 are replicates of experiments performed using non-supercoiled linear DNA *in vitro* at 1mM NTPs. Analysis of the data derived from these samples (Table S1) is presented in Figures 2, 3B (X-axis), 3D (X-axis), 3E, 3F, 5C, 7A, S5, and S6A.
(ii) VV866 and VV867 are replicates of experiments performed using negatively supercoiled DNA *in vitro* at 1mM NTPs. Analysis of the data derived from these samples (Table S2) is presented in Figures 3A, 3B (Y-axis), 3E, 3F, 5C, 7B, S3A, S5, and S6B.
(iii) VV868, Vv869, and VV870 are replicates of experiments performed using negatively supercoiled DNA *in vivo*. Analysis of the data derived from these samples (Table S3) is presented in Figures 3C, 3D (Y-axis), Figure 3E, 5C, 7C, S3B, S5, and S6C

(iv) VV852 and VV853 are replicates of experiments performed using non-supercoiled linear DNA *in vitro* at saturating NTPs. Analysis of the data derived from these samples (Table S4) is presented in Figures 4A, 4C, 4E (Y-axis), 4F, 6A and 6C.
(v) VV856 and VV857 are replicates of experiments performed using non-supercoiled linear DNA *in vitro* at non-saturating NTPs. Analysis of the data derived from these samples (Table S5) is presented in Figures 4B, 4D, 4E (X-axis), 4F, 5C, 6B and 6C.
(vi) VV886 and VV887 are replicates of experiments performed using non-supercoiled relaxed circular DNA *in vitro* at 1mM NTPs. Analysis of the data derived from these samples (Table S6) is presented in Figures S4A, S4C (X-axis), S4D, and S4E.
(vii) VV889 and VV890 are replicates of experiments performed using negatively supercoiled DNA *in vitro* at 1mM NTPs performed in parallel with experiments done using non-supercoiled relaxed circular DNA. Analysis of the data derived from these samples (Table S7) is presented in Figures S4B, S4C (Y-axis), S4D, and S4E.

Data analysis

Source code and documentation for analysis of DNA templates and 5′ RNA seq libraries are provided at: https://github.com/NickelsLabRutgers/MASTER-Data-Analysis.

DNA template analysis. High-throughput sequencing of template DNA was used to associate the 7-bp randomized sequence in the region of interest with a corresponding second 15-bp randomized sequence that serves as its barcode. The linear DNA templates used for *in vitro* transcription reactions can be sequenced directly because they contain sequences required for compatibility with the Illumina HiSeq at both ends of the molecule (i.e. sequences that enable the DNA to bind to complementary sequences attached to the Illumina flow cell during bridge amplification). Plasmid-borne DNA templates cannot be sequenced directly. Therefore to assign the barcodes on plasmid-borne templates by use of high-throughput sequencing we use the plasmid as template in an emulsion PCR reaction using Illumina RP1 primer and Illumina RPI index primer. The ePCR product generated from this procedure carries the sequence of the promoter, the sequence of the 7-bp region of interest, and the sequence of the 15-bp index region.

We analyzed the first 80 bases of each sequencing read to identify those that were a perfect match to the conserved regions of the MASTER-*lacCONS-N7* template sequence:

aggc**ttgaca**ctttatgcttcggctcg**tataat**gtgnnnnnnngtgagcggataacaatnnnnnnnnnnnnnnnntggaattctcgggtg ccaagg (the -35 element and -10 element of p*lacCONS* are in bold; the 7-bp region of interest and the 15-bp barcode region are underlined). From these reads we identified 15-bp barcodes that could be uniquely assigned to one of the 16,384 potential sequences in the randomized region of interest. Specifically, we identified 15-bp barcodes for which ≥95% of the sequencing reads containing this barcode contained an identical 7-bp region of interest. These 15-bp barcodes were used to associate RNA reads with their template of origin.

5′ RNA-seq analysis: Reads that contain a perfect match to the sequence downstream of position 10 and a perfect match to 5 bases of sequence downstream of the 15-bp barcode were identified. The identity of the 15-bp barcode was used to determine the identity of the bases at position 4-10 of the *lacCONS* template (by means of the associations established in the above analysis of the DNA templates).

Sequences derived from the RNA 5′ end of reads that were perfect matches to the sequence of the template from which they were derived were used to identify the TSS position. Sequences derived from the RNA 5′ end of reads that carried one or more mismatches from the DNA template were not considered in the analysis of TSS position, but were used for analysis of productive slippage synthesis and for analysis of transcript yields. We note that sequences derived from the RNA 5′ end begin at the 16th bp of each read because the 5′ adaptor used to generate the cDNAs carries 15 randomized bases at the 3′ end to improve ligation efficiency and provide a "digital tag" to individual transcripts. These 15-base digital tags are used to identify cDNAs that are preferentially amplified during PCR and count these species as only a single read during analysis.

Productive slippage synthesis analysis: MASTER

For each TSS-region sequence with the patterns specified below, read counts of matched RNAs (i.e. reads derived from RNAs generated by standard synthesis) and read counts of slippage RNAs were determined using the data in Tables S1-S7. The slippage rate (% slippage) was calculated using the read counts derived from all TSS-region sequences with the specified TSS-region sequence pattern.

$$100 * [\sum slippage\ read\ count / (\sum slippage\ read\ count + \sum matched\ read\ count).]$$

Productive slippage synthesis analysis: E. coli chromosome

Analysis of slippage from *E. coli* chromosomal promoters was performed using data files SRR1173966, SRR1173969, SRR1173970, SRR1173973, SRR1173974, SRR1173978, SRR1173979, SRR1173980, SRR1173982, SRR1173985, and SRR1173986 from (Thomason et al., 2015) available in the NIH SRA database. Matched RNA reads and slippage reads emanating from the 14,868 annotated TSS reported in (Thomason et al., 2015) were identified in the following manner. First, for each of the 14,868 TSS, the DNA sequences 8 bp upstream to 29 bp downstream (positions -8 to +30) were extracted, and the sequence from -3 to +4 was defined as a TSS-region sequence. Next, we identified reads with a quality score >25 for the first 38 bases. Using these reads we tested substrings in $\{\langle RNA[i,j],(i,j)\rangle \big| (i,j) \in \cup_{i=1}^{12} \cup_{j=26+i}^{38} \{(i,j)\}\}$ for matching to TSS-region sequence substrings in $\{\langle DNA[k,38],DNA\rangle | DNA \in \{TSS\ sequence\}, k = 1,2,\dots,12\}$. Reads that matched a TSS-region substring were analyzed for matched RNA reads and slippage RNA reads as described above using TSS-region sequences with the specified patterns. We note that for the analysis of slippage from *E. coli* chromosomal promoters we considered slippage from homopolymeric repeat sequences that start at an annotated TSS (T_n, A_n, G_n, and C_n, where n > 1 and where the sequence starts at an annotated TSS; Figure S4B, top) or homopolymeric repeat sequences that start one base downstream of a TSS (T_n, A_n, G_n, and C_n, where n > 1 and where the sequence starts one base downstream of an annotated TSS; Figure S5B, bottom). The slippage rate (% slippage) was calculated using the read counts derived from all TSS-region sequences with the specified TSS-region sequence pattern.

$100 * [\sum slippage\ read\ count / (\sum slippage\ read\ count + \sum matched\ read\ count).]$

Productive slippage synthesis analysis: patterns and sequences of standard reads and slippage reads.

homopolymeric repeats that start at the TSS		
TSS-region*	Sequence of standard RNA	Sequence of slippage RNAs
NNAANNN	AANNN	A/AA/AAA/... + AANNN
NNCCNNN	CCNNN	C/CC/CCC/... + CCNNN
NNGGNNN	GGNNN	G/GG/GGG/... + GGNNN
NNTTNNN	TTNNN	T/TT/TTT/... + TTNNN
NNAAANN	AAANN	A/AA/AAA/... + AAANN
NNCCCNN	CCCNN	C/CC/CCC/... + CCCNN
NNGGGNN	GGGNN	G/GG/GGG/... + GGGNN
NNTTTNN	TTTNN	T/TT/TTT/... + TTTNN
NNAAAAN	AAAAN	A/AA/AAA/... + AAAAN

NNCCCCN	CCCCN	C/CC/CCC/... + CCCCN
NNGGGGN	GGGGN	G/GG/GGG/... + GGGGN
NNTTTTN	TTTTN	T/TT/TTT/... + TTTTN
NNAAAAA	AAAAA	A/AA/AAA/... + AAAAA
NNCCCCC	CCCCC	C/CC/CCC/... + CCCCC
NNGGGGG	GGGGG	G/GG/GGG/... + GGGGG
NNTTTTT	TTTTT	T/TT/TTT/... + TTTTT
NNNAANN	AANN	A/AA/AAA/... + AANN
NNNCCNN	CCNN	C/CC/CCC/... + CCNN
NNNGGNN	GGNN	G/GG/GGG/... + GGNN
NNNTTNN	TTNN	T/TT/TTT/... + TTNN
NNNAAAN	AAAN	A/AA/AAA/... + AAAN
NNNCCCN	CCCN	C/CC/CCC/... + CCCN
NNNGGGN	GGGN	G/GG/GGG/... + GGGN
NNNTTTN	TTTN	T/TT/TTT/... + TTTN
NNNAAAA	AAAA	A/AA/AAA/... + AAAA
NNNCCCC	CCCC	C/CC/CCC/... + CCCC
NNNGGGG	GGGG	G/GG/GGG/... + GGGG
NNNTTTT	TTTT	T/TT/TTT/... + TTTT
NNNNAAN	AAN	A/AA/AAA/... + AAN
NNNNCCN	CCN	C/CC/CCC/... + CCN
NNNNGGN	GGN	G/GG/GGG/... + GGN
NNNNTTN	TTN	T/TT/TTT/... + TTN
NNNNAAA	AAA	A/AA/AAA/... + AAA
NNNNCCC	CCC	C/CC/CCC/... + CCC
NNNNGGG	GGG	G/GG/GGG/... + GGG
NNNNTTT	TTT	T/TT/TTT/... + TTT
NNNNNAA	AA	A/AA/AAA/... + AA
NNNNNCC	CC	C/CC/CCC/... + CC
NNNNNGG	GG	G/GG/GGG/... + GG
NNNNNTT	TT	T/TT/TTT/... + TT
homopolymeric repeats that start 1bp downstream of the TSS		
TSS-region*	Sequence of standard RNA	Sequence of slippage RNAs
NNNAANN	NAANN	NA/NAA/NAAA... + AANN
NNNCCNN	NCCNN	NC/NCC/NCCC... + CCNN
NNNGGNN	NGGNN	NG/NGG/NGGG... + GGNN
NNNTTNN	NTTNN	NT/NTT/NTTT... + TTNN
NNNAAAN	NAAAN	NA/NAA/NAAA... + AAAN
NNNCCCN	NCCCN	NC/NCC/NCCC... + CCCN
NNNGGGN	NGGGN	NG/NGG/NGGG... + GGGN
NNNTTTN	NTTTN	NT/NTT/NTTT... + TTTN
NNNAAAA	NAAAA	NA/NAA/NAAA... + AAAA

NNNCCCC	NCCCC	NC/NCC/NCCC... + CCCC
NNNGGGG	NGGGG	NG/NGG/NGGG... + GGGG
NNNTTTT	NTTTT	NT/NTT/NTTT... + TTTT
NNNNAAN	NAAN	NA/NAA/NAAA... + AAN
NNNNCCN	NCCN	NC/NCC/NCCC... + CCN
NNNNGGN	NGGN	NG/NGG/NGGG... + GGN
NNNNTTN	NTTN	NT/NTT/NTTT... + TTN
NNNNAAA	NAAA	NA/NAA/NAAA... + AAA
NNNNCCC	NCCC	NC/NCC/NCCC... + CCC
NNNNGGG	NGGG	NG/NGG/NGGG... + GGG
NNNNTTT	NTTT	NT/NTT/NTTT... + TTT
NNNNNAA	NAA	NA/NAA/NAAA... + AA
NNNNNCC	NCC	NC/NCC/NCCC... + CC
NNNNNGG	NGG	NG/NGG/NGGG... + GG
NNNNNTT	NTT	NT/NTT/NTTT... + TT
non-homopolymeric repeats at the TSS		
TSS-region	Sequence of standard RNA	Sequence of slippage RNAs
NNNACNN	ACNN	AC/ACAC/ACACAC/... + ACNN
NNNAGNN	AGNN	AG/AGAG/AGAGAG/... + AGNN
NNNATNN	ATNN	AT/ATAT/ATATAT/... + ATNN
NNNCANN	CANN	CA/CACA/CACACA/... + CANN
NNNCGNN	CGNN	CG/CGCG/CGCGCG/... + CGNN
NNNCTNN	CTNN	CT/CTCT/CTCTCT/... + CTNN
NNNGANN	GANN	GA/GAGA/GAGAGA/... + GANN
NNNGCNN	GCNN	GC/GCGC/GCGCGC/... + GCNN
NNNGTNN	GTNN	GT/GTGT/GTGTGT/... + GTNN
NNNTANN	TANN	TA/TATA/TATATA/... + TANN
NNNTCNN	TCNN	TC/TCTC/TCTCTC/... + TCNN
NNNTGNN	TGNN	TG/TGTG/TGTGTG/... + TGNN

* Bases upstream and downstream of the homopolymeric repeat sequences are different from the bases that comprise the repeat sequence.

Relative expression analysis. The relative expression of full-length transcripts from each sequence variant were defined as the ratio of the total number of 5′ RNA-seq reads generated from a given TSS-region sequence variant to the relative number of DNA templates within the library that contained the TSS-region sequence variant. For experiments performed using plasmid-borne libraries *in vitro* and *in vivo* we determined the relative number of DNA templates by high-throughput sequencing of an emulsion PCR product that was generated using the

plasmid-borne library as template. For experiments performed using a linear template, we determined the relative number of DNA templates by directly sequencing the template.

Statistical analysis (Figure S2). Positional preference of TSS selection was analyzed using the Tukey Honest Significant Differences (Tukey HSD) test (Figure S2A). Tukey HSD test was performed on the mean percentages of matched RNA reads emanating from positions 6-10 of TSS-region sequences with ≥25 total matched RNA reads in each of the seven experimental data sets (Tables S1-S7). The Shapiro-Wilk test of normality indicated that the mean percentages of each position are drawn from normal distribution. Results indicate that the mean percentage of reads derived from matched RNA reads emanating from position 7 is significantly higher than that of all other positions ($p < 0.001$; Figure S2A), the mean percentage of matched RNA reads emanating from position 8 is significantly higher than that of positions 6, 9, and 10 ($p < 0.001$; Figure S2A), and the mean percentage of matched RNA reads emanating from positions 6 and 9 have no significant difference between each other (Figure S2A). All statistical tests were performed in R (http://www.R-project.org/).

Sequence preference of TSS selection (Figure 2, 4 and S4) was analyzed using the paired Wilcoxon signed-rank test (Figure S2B-D). The test was performed on the mean percentages of matched RNA reads emanating from positions 6-10 in each of the seven experimental data sets (Tables S1-S7). Results indicate that for each position the mean percentage of reads emanating from a pyrimidine (R) are significantly higher than the mean percentage of reads that emanate from a purine (Y; $p < 0.001$; Figure S2B). In addition, for each position the mean percentage of reads emanating from $Y_{TSS-1}R_{TSS}$ is significantly larger than the mean percentage of reads that emanate from $R_{TSS-1}R_{TSS}$ ($p < 0.001$; Figure S2D). Results from analysis on specific nucleotides further show that the mean percentage of reads that emanate from a G is significantly larger than the mean percentage of reads emanating from an A, C, or T ($p < 0.001$, adjusted by Bonferroni correction; Figure S2C), and the mean percentage of reads emanating from an A is significantly larger than the mean percentage of reads that emanate from a C or a T ($p < 0.001$, adjusted by Bonferroni correction; Figure S2C). Results further indicate that the mean percentages of reads emanating from a C or a T have no significant difference between each other (Figure S2C).

Sensitivity of TSS selection to DNA topology (Figures 3 and S4) was analyzed using paired Wilcoxon signed-rank test. Results indicate that the mean TSS observed in experiments performed using supercoiled DNA templates *in vitro* (Tables S2 and S7) is significantly larger

than the mean TSS observed in experiments performed using non-supercoiled DNA templates *in vitro* (Tables S1 and S6; $p < 0.001$) and the mean TSS observed in experiments performed using supercoiled DNA templates *in vivo* (Table S3) is significantly larger than the mean TSS observed in experiments performed using non-supercoiled DNA templates *in vitro* (Tables S1 and S6; $p < 0.001$).

The sensitivity of specific TSS-region sequences to alterations in DNA topology and alterations in NTP concentrations were analyzed using paired Wilcoxon signed-rank test (Figure S2E). Results indicate that TSS region sequences containing an R at position 7 or 8, an RR at positions 7/8, or a YR at positions 6/7 have significantly lower changes in the mean TSS in response to alterations in DNA topology and in response to alterations in NTP concentrations ($p < 0.001$; Figure S2E). Results further indicate that sequences containing a Y at position 7 or 8, a YY at positions 7/8, an RY at positions 6/7, an RYYR at positions 6-9, or an RYYYR at positions 6-10 have significantly higher changes in the mean TSS in response to alterations in DNA topology and in response to alterations in NTP concentrations ($p < 0.001$; Figure S2E).

Correlation between the precision of TSS-selection and transcript yield was analyzed using Wilcoxon signed-rank test (Figure S2F). The test compared the TSS variance of TSS-region sequences with the highest 1% relative expression with the TSS variance of the remaining 99% of TSS-region sequences. Results indicate that TSS-region sequences with the highest 1% relative expression have significantly less TSS variance than the remaining 99% of TSS-region sequences ($p < 0.001$; Figure S2F). In addition, TSS-region sequences with the highest 1% relative expression have significantly less minimum distance from position 7 or 8 than the remaining 99% of TSS-region sequences ($p < 0.001$).

Reproducibility of experimental replicates was analyzed using hierarchical clustering and Wilcoxon signed-rank test (Figure S2G). Hierarchical clustering was performed on the Euclidean distances between the percentages of reads derived from matched RNAs that emanate from positions 4-10 of TSS-region sequences with ≥25 total matched RNA reads in each of the seven experimental data sets (Tables S1-S7) using equation S1.

$$\boldsymbol{D} = \|\boldsymbol{p} - \boldsymbol{q}\| = \sqrt{\sum_{\boldsymbol{i=1}}^{\boldsymbol{m*n}} (\boldsymbol{p_i} - \boldsymbol{q_i})^{\boldsymbol{2}}} \quad \text{(S1)}$$

In equation S1, p and q are the vectors of percentages of reads emanating from positions 4-10 for all TSS-region sequences, m is the total number of TSS-region sequences with ≥25 total matched RNA reads in each of the seven experimental data sets, and $n = 7$ (i.e. the number of positions being considered). Results indicate that replicates are more similar to each other than to non-replicates. Wilcoxon signed-rank test on the distances between replicates or non-replicates further indicates that the distances between non-replicates are significantly larger than those between replicates ($p < 0.001$).

Similarity between the results of experiments performed using supercoiled DNA templates *in vitro* (Tables S2 and S7) and the results of experiments performed using supercoiled DNA templates *in vivo* (Table S3) was analyzed using hierarchical clustering and Wilcoxon signed-rank test. Hierarchical clustering results of reproducibility analysis indicate that results obtained using supercoiled DNA templates *in vivo* are more similar to results obtained using supercoiled DNA templates *in vitro* than to results obtained using non-supercoiled linear DNA templates *in vitro*. Wilcoxon signed-rank test was performed on the distances calculated using equation S1 between results obtained using linear templates *in vitro*, supercoiled templates *in vitro*, and supercoiled templates *in vivo*. Results indicate that the distance between results obtained using supercoiled DNA templates *in vivo* and results obtained using supercoiled DNA templates *in vitro* is significantly lower than (i) the distance between results obtained using supercoiled DNA templates *in vitro* and non-supercoiled DNA templates *in vitro* and (ii) the distance between results obtained using supercoiled DNA templates *in vivo* and non-supercoiled DNA templates *in vitro* ($p < 0.001$).

Data deposition. Raw reads have been deposited in the NIH/NCBI Sequence Read Archive under the study accession number SRP057850.

SUPPLEMENTAL REFERENCES

Goldman, S.R., Sharp, J.S., Vvedenskaya, I.O., Livny, J., Dove, S.L., and Nickels, B.E. (2011). NanoRNAs prime transcription initiation in vivo. Mol. Cell *42*, 817-825.

Jamroze, A., Perugino, G., Valenti, A., Rashid, N., Rossi, M., Akhtar, M., and Ciaramella, M. (2014). The reverse gyrase from *Pyrobaculum calidifontis*, a novel extremely thermophilic DNA topoisomerase endowed with DNA unwinding and annealing activities. J. Biol. Chem. *289*, 3231-3243.

Napoli, A., Valenti, A., Salerno, V., Nadal, M., Garnier, F., Rossi, M., and Ciaramella, M. (2005). Functional interaction of reverse gyrase with single-strand binding protein of the archaeon Sulfolobus. Nucleic Acids Res. *33*, 564-576.

Vvedenskaya, I.O., Vahedian-Movahed, H., Bird, J.G., Knoblauch, J.G., Goldman, S.R., Zhang, Y., Ebright, R.H., and Nickels, B.E. (2014). Interactions between RNA polymerase and the "core recognition element" counteract pausing. Science *344*, 1285-1289.